PRODUCTION OF PROTEINS USING HOMOLOGOUS RECOMBINATION

INTRODUCTION

Technical Field

The field of this invention is the expression of mammalian proteins.

Background

The discoveries of restriction enzymes, cloning, sequencing, reverse transcriptase, and monoclonal antibodies has resulted in extraordinary capabilities in isolating, identifying, and manipulating nucleic acid sequences. As a result of these capabilities, numerous genes and their transcriptional control elements have been identified and manipulated. The genes have been used for producing large amounts of a desired protein in heterologous hosts (bacterial and eukaryotic host cell systems).

In many cases, the process of obtaining coding sequences and eliciting their expression has been a long and arduous one. The identification of the coding sequence, either cDNA or genomic DNA, has frequently involved the construction of libraries, identification of fragments of the open reading frame, examining the flanking sequences, and the like. In mammalian genes where introns are frequently encountered, in many instances, the coding region has been only a small fraction of the total nucleic acid associated with the gene. In other cases, pseudogenes or multi-membered

gene families have obscured the ability to isolate a particular gene of interest. Nevertheless, as techniques have improved, there has been a continuous parade of successful identifications and isolation of genes of interest.

In many situations one is primarily interested in a source of the protein product. The cell type in the body which produces the protein is frequently an inadequate source, since the protein may be produced in low amounts, the protein may only be produced in a differentiated host cell which is only difficultly grown in culture, or the host cell, particularly a human cell, is not economic or efficient in a culture process for production of the product. There is, therefore, significant interest in developing alternative techniques for producing proteins of interest in culture with cells which provide for economic and efficient production of the desired protein and, when possible, appropriate processing of the protein product.

Relevant Literature

Mansour et al., Nature, 336:348-352, (1988), describe a general strategy for targeting mutations to non-selectable genes. Weidle et al., Gene, 66:193-203, (1988), describe amplification of tissue-type plasminogen activator with a DHFR gene and loss of amplification in the absence of selective pressure. Murnane and Yezzi, Somatic Cell and Molecular Genetics, 14:273-286, (1988), describe transformation of a human cell line with an integrated selectable gene marker lacking a transcriptional promoter, with tandem duplication and amplification of the gene marker. Thomas and Capecchi, Cell, 51:503-512, (1987), describe site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Song et al., Proc. Natl. Acad. Sci. USA, 84:6820-6824, (1987), describe

homologous recombination in human cells by a two staged integration. Liskay et al., "Homologous Recombination Between Repeated Chromosomal Sequences in Mouse Cells," Cold Spring Harbor, Symp. Quant. Biol. 49:183-189, (1984), describe integration of two different mutations of the same gene and homologous recombination between the mutant genes. Rubnitz and Subramani, Mol. and Cell. Biol. 4:2253-2258, (1984), describe the minimum amount of homology required for homologous recombination in mammalian cells. Kim and Smithies, Nucl. Acids. Res. 16:8887-8903, (1988), describe an assay for homologous recombination using the polymerase chain reaction.

SUMMARY OF THE INVENTION

Expression of mammalian proteins of interest is achieved by employing homologous recombination for integration of an amplifiable gene and other regulatory sequences in proximity to the gene of interest without interruption of the production of a proper transcript. The region comprising the amplifiable gene and the gene of interest may be amplified, the genome fragmented and directly or indirectly transferred to an expression host for expression of the target protein. If not previously amplified, the target region is then amplified, and the cell population screened for cells producing the target protein. Cells which produce the target protein at high and stable levels are expanded and used for expression of the target protein.

DESCRIPTION OF SPECIFIC EMBODIMENTS

Methods and compositions are provided for production of mammalian proteins of interest in culture. The method employs homologous recombination in a host cell for integrating an amplifiable gene in the vicinity of a target gene, which target gene encodes the protein of interest. The region comprising

both the amplifiable gene and target gene will be referred to as the amplifiable region. The resulting transformed primary cells may now be subjected to conditions which select for amplification, or the amplification may be performed subsequently. "Transform" includes transform, transfect, transduce, conjugation, fusion, electroporation or any other technique for introducing DNA into a viable cell. The chromosomes or DNA of the transformed cells are then used to transfer the amplifiable region into the genome of secondary expression host cells, where the target region, if not previously amplified, is amplified. The resulting cell lines are screened for production of the target protein and secondary cell lines selected for desired levels of production, which cells may be expanded and used for production of the desired protein in culture.

The primary cell may be any mammalian cell of interest, particularly mammalian cells which do not grow readily in culture, more particularly primate cells, especially human cells, where the human cells may be normal cells or neoplastic cells, particularly normal cells. Various cell types may be employed as the primary cells, including fibroblasts, particularly diploid skin fibroblasts, lymphocytes, epithelial cells, neurons, endothelial cells, or other somatic cells, or germ cells. Of particular interest are skin fibroblasts, which can be readily propagated to provide for large numbers of normal cells. These cells may or may not be expressing the gene of interest. In those instances where the target gene is inducible or only expressed in certain differentiated cells, one may select cells in which the target gene is expressed, which may require immortalized cells capable of growth in culture.

A number of amplifiable genes exist, where by appropriate use of a selection agent, a gene integrated

in the genome will be amplified with adjacent flanking DNA. Amplifiable genes include dihydrofolate reductase, metallothionein-I and -II, preferably primate metallothionein genes, adenosine deaminase, ornithine decarboxylase, etc. The amplifiable gene will have transcriptional signals which are functional in the secondary or expression host and desirably be functional in the primary host, particularly where amplification is employed in the primary host or the amplifiable gene is used as a marker.

The target genes may be any gene of interest, there already having been a large number of proteins of interest identified and isolated with continual additions to the list. Proteins of interest include cytokines, such as interleukins 1-7; growth factors such as EGF, FGF, PDGF, and TGF; somatotropins; growth hormones; colony stimulating factors, such as G-, M-, and GM-CSF; erythropoietin; plasminogen activators, such as tissue and urine; enzymes, such as superoxide dismutase; interferons; T-cell receptors; surface membrane proteins; insulin; lipoproteins; α_1-antitrypsin; CD proteins, such as CD3, 4, 8, 19; clotting factors, e.g., Factor VIIIc and von Willebrands factor; anticlotting factors, such as Protein C; atrial naturetic factor, tumor necrosis factor; etc.

For homologous recombination, constructs will be prepared where the amplifiable gene will be flanked on one or both sides with DNA homologous with the DNA of the target region. The homologous DNA will generally be within 100 kb, usually 50 kb, preferably about 25 kb, of the transcribed region of the target gene, more preferably within 2 kb of the target gene. By gene is intended the coding region and those sequences required for transcription of a mature mRNA. The homologous DNA may include the 5'-upstream region comprising any enhancer sequences,

transcriptional initiation sequences, the region 5' of these sequences, or the like. The homologous region may include a portion of the coding region, where the coding region may be comprised only of an open reading frame or combination of exons and introns. The homologous region may comprise all or a portion of an intron, where all or a portion of one or more exons may also be present. Alternatively, the homologous region may comprise the 3'-region, so as to comprise all or a portion of the transcription termination region, or the region 3' of this region. The homologous regions may extend over all or a portion of the target gene or be outside the target gene comprising all or a portion of the transcriptional regulatory regions and/or the structural gene. For the most part, the homologous sequence will be joined to the amplifiable gene, proximally or distally, usually a sequence other than the wild-type sequence normally associated with the target gene will be used to separate the homologous sequence from the amplifiable gene on at least one side of the amplifiable gene. Some portion of the sequence will be the 5' or 3' sequence associated with the amplifiable gene, as a result of the manipulations associated with the amplifiable gene.

The homologous regions flanking the amplifiable gene need not be identical to the target region, where in vitro mutagenesis is desired. For example, one may wish to change the transcriptional initiation region for the target gene, so that a portion of the homologous region might comprise nucleotides different from the wild-type 5'-region of the target gene. Alternatively, one could provide for insertion of a transcriptional initiation region different from the wild-type initiation region between the wild-type initiation region and the structural gene. Similarly, one might wish to introduce various mutations into the structural gene, so that the

homologous region would comprise mismatches, resulting in a change in the encoded protein. For example, a signal leader sequence would be introduced in proper reading frame with the target gene to provide for secretion of the target protein expression product. Alternatively, one might change the 3' region, e.g., untranslated region, polyadenylation site, etc. of the target gene. Therefore, by homologous recombination, one can provide for maintaining the integrity of the target gene, so as to express the wild-type protein under the transcriptional regulation of the wild-type promoter or one may provide for a change in transcriptional regulation, processing or sequence of the target gene. In some instances, one may wish to introduce an enhancer in relation to the transcriptional initiation region, which can be provided by, for example, integration of the amplifiable gene associated with the enhancer in a region upstream from the transcriptional initiation regulatory region or in an intron or even downstream from the target gene.

In order to prepare the subject constructs, it will be necessary to know the sequence which is targeted for homologous recombination. While it is reported that a sequence of 14 bases complementary to a sequence in a genome may provide for homologous recombination, normally the individual flanking sequences will be at least about 150 bp, and may be 10 kb or more, usually not more than about 5 kb. The size of the flanking regions will be determined by the size of the known sequence, the number of sequences in the genome which may have homology to the site for integration, whether mutagenesis is involved and the extent of separation of the regions for mutagenesis, the particular site for integration, or the like.

The integrating constructs may be prepared in accordance with conventional ways, where sequences may

be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to in vitro mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like. Usually the construct will be carried on a cloning vector comprising a replication system functional in a prokaryotic host, e.g., E. coli, and a marker for selection, e.g., biocide resistance, complementation to an auxotrophic host, etc. Other functional groups may also be present, such as polylinkers, for ease of introduction and excision of the construct or portions thereof, or the like. A large number of cloning vectors are available such as pBR322, the pUC series, etc.

Once the construct is prepared, it may then be used for homologous recombination in the primary cell target. Various techniques may be employed for integrating the construct into the genome of the primary cell without being joined to a replication system functional in the primary host. See for example, U.S. Patent No. ~~4,319,216~~ 4,399,216, as well as the references cited in the Relevant Literature section. Alternatively, the construct may be inserted into an appropriate vector, usually having a viral replication system, such as SV40, bovine papilloma virus, adenovirus, or the like, where the vector may also have a selectable marker for identifying transfected cells. Selectable markers include the neo gene, allowing for selection with G418, the herpes tk gene for selection with HAT medium, gpt gene with mycophenolic acid, complementation of an auxotrophic host, etc.

The vector may or may not be capable of stable maintenance in the host. Where the vector is capable of stable maintenance, the cells will be screened for homologous integration of the vector into the genome of

the host, where various techniques for curing the cells may be employed. Where the vector is not capable of stable maintenance, for example, where a temperature sensitive replication system is employed, one may change the temperature from the permissive temperature to the non-permissive temperature, so that the cells may be cured of the vector. In this case, only those cells having integration of the construct comprising the amplifiable gene and, when present, the selectable marker, will be able to survive selection.

Where a selectable marker is present, one may select for the presence of the construct by means of the selectable marker. Where the selectable marker is not present, one may select for the presence of the construct by the amplifiable gene. For the neo gene or the herpes tk gene, one could employ a medium for growth of the transformants of about 0.1-1 mg/ml of G418 or HAT medium respectively. Where DHFR is the amplifiable gene, the selective medium may include from about 0.01-0.25 μM of methotrexate.

In carrying out the homologous recombination, the DNA will be introduced into the primary cells. Techniques which may be used include calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, electroporation, bacterial protoplast fusion with intact cells, transfection, polycations, e.g., polybrene, polyornithine, etc., or the like. The DNA may be single or double stranded DNA. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology (1989), in press, and Mansour et al., Nature, 336:348-352, (1988).

Upstream and/or downstream from the target region construct may be a gene which provides for identification of whether a double crossover has occurred. For this purpose, the herpes simplex virus thymidine kinase gene may be employed since the presence of the thymidine kinase gene may be detected

by the use of nucleoside analogs, such as acyclovir or gancyclovir, for their cytotoxic effects on cells that contain a functional HSV-tk gene. The absence of sensitivity to these nucleoside analogs indicates the absence of the thymidine kinase and, therefore, where homologous recombination has occurred, that a double crossover event has also occurred.

The presence of the marker gene as evidenced by resistance to a biocide or growth in a medium which selects for the presence of the marker gene, establishes the presence and integration of the target construct into the host genome. No further selection need be made at this time, since the selection will be made in the secondary expression host, where expression of the amplified target gene may be detected. If one wishes, one can determine whether homologous recombination has occurred by employing PCR and sequencing the resulting amplified DNA sequences. If desired, amplification may be performed at this time by stressing the primary cells with the appropriate amplifying reagent, so that multi-copies of the target gene are obtained. Alternatively, amplification may await transfer to the secondary cell expression host.

High molecular weight DNA, greater than about 20kb, preferably greater than about 50kb DNA or preferably metaphase chromosomes are prepared from the primary recipient cell strain having the appropriate integration of the amplification vector. Preparation and isolation techniques are described by Nelson and Housman, In Gene Transfer (ed. R. Kucherlapati) Plenum Press, 1986. The DNA may then be introduced in the same manner as described above into the secondary host expression cells, using the same or different techniques than employed for the primary cells.

Various mammalian expression hosts are available and may be employed. These hosts include CHO cells, monkey kidney cells, C127 mouse fibroblasts, 3T3 mouse cells,

Vero cells, etc. Desirably the hosts will have a negative background for the amplifiable gene or a gene which is substantially less responsive to the amplifying agent.

The transformed cells are grown in selective medium containing about 0.01-0.5 μM methotrexate and, where another marker is present, e.g., the neo gene, the medium may contain from about 0.1-1 mg/ml G418. The resistant colonies are isolated and may then be analyzed for the presence of the construct in juxtaposition to the target gene. This may be as a result of detection of expression of the target gene product, where there will normally be a negative background for the target gene product, use of PCR, Southern hybridization, or the like.

The cells containing the construct are then expanded and subjected to selection and amplification with media containing progressively higher concentrations of the amplifying reagent, for example, 0.5-200 μM of methotrexate for the DHFR gene, and may be analyzed at each selection step for production of the target product.

The various clones may then be screened for optimum stable production of the target product and these clones may then be expanded and used commercially for production in culture. In this manner, high yields of a product may be obtained, without the necessity of isolating the message and doing the various manipulations associated with genetic engineering or isolating the genomic gene, where very large genes can be a major research and development effort.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

Cells

Normal human diploid skin fibroblasts, ("primary recipient") are propagated in EEMEM medium supplemented with 20% fetal calf serum. Dihydrofolate reductase (DHFR) deficient Chinese hamster ovary (CHO) DUKX-B11 cells (Urlaub and Chasin, Proc. Natl. Acad. Sci. USA 77:4216-4220 (1980)) ("secondary recipient") are propagated in alpha-medium supplemented with 10% dialyzed fetal bovine serum.

DNA Vector

The amplification vector is constructed from pUC19 (Yanisch-Perron et al., Gene 33:103-119 (1985)). A 1.8 kb HaeII fragment containing a hygromycin B phosphotransferase gene (hph) driven by the herpes simplex virus thymidine kinase (HSV tk) promoter is isolated from pHyg (Sugden et al., Mol. Cell. Biol. 5:410-413 (1985)) by digestion with HaeII and gel electrophoresis. Synthetic adaptors are added onto this fragment to convert the HaeII ends into HindIII ends and the resulting fragment is joined to pUC19 digested with HindIII. The resulting plasmid pUCH contains the hygromycin cassette such that transcription of hph and beta-lactamase are in the opposite orientation. A 1.3 kb SalI fragment containing a DHFR gene driven by SV40 transcriptional signals is isolated from pTND (Connors et al., DNA 7:651-661 (1988)) by digestion with SalI and gel electrophoresis. This fragment is ligated to pUCH digested with SalI. The resulting plasmid pUCD contains the DHFR cassette such that DHFR and hph are transcribed in the same direction. A 1.76 kb BamHI fragment from the phage F15 (Friezner Degen et al., J. Biol. Chem. 261:6972-6985 (1986)) which contains 1.45 kb of DNA flanking the transcriptional start of human

tissue plasminogen activator (t-PA) in addition to the first exon and part of the first intron is isolated by gel electrophoresis after BamHI digestion. This fragment is joined to pUCD following digestion of the latter with BamHI. The resulting plasmid pUCG has the promoter of the t-PA fragment oriented opposite to that of the DHFR cassette. The t-PA fragment contains a single NcoI site, which is not unique to pUCG. A partial NcoI digest is carried out and a NotI linker is inserted. The resulting plasmid pCG contains a unique NotI site in the t-PA fragment which allows the plasmid to be linearized prior to transformation of the primary human diploid fibroblasts in order to increase the frequency of homologous recombination (Kucherlapati et al., Proc. Natl. Acad. Sci. USA 81:3153-3157 (1984)).

Preparation of Primary Recipients

The plasmid pCG linearized with NotI is introduced into the primary recipients by electroporation employing DNA at 10nM. The resulting cells are then grown in selective medium (EEMEM with 200 µg/ml hygromycin B). Resistant colonies are isolated and analyzed by PCR (Kim and Smithies, Nucleic Acids Res. 16:8887-8903 (1988)) using as primers the sequences GCGGCCTCGGCCTCTGCATA and CATCTCCCCTCTGGAGTGGA to distinguish homologous integrants from random ones. Amplification of cellular DNA by PCR using these two primers yields a fragment of 1.9 kb only when DNA from correctly targeted cells is present. Cells comprising the DHFR gene integrated into the t-PA region are expanded and used as a source of genetic material for preparation of secondary recipients.

Preparation of Secondary Recipients

Metaphase chromosomes are prepared (Nelson et al., J. Mol. Appl. Genet. 2:563-577 (1984)) from recipients demonstrating homologous recombination with

the DHFR and are then transformed in DHFR-deficient CHO cells by calcium phosphate mediated gene transfer (Nelson et al., J. Mol. Appl. Genet. 2:563-577 (1984)). The cells are then grown in selective medium (alpha-medium containing 200 µg/ml hygromycin B). Resistant colonies are isolated and analyzed for expression of human t-PA (Kaufman et al., Mol. Cell. Biol. 5:1750-1759 (1985)). The cell clones are then grown in selective medium containing progressively higher concentrations of methotrexate (.02-80 µM, with steps of 4-fold increases in concentration). After this amplification procedure, the cells are harvested and the human t-PA is analyzed employing an ELISA assay with a monoclonal antibody specific for t-PA (Weidle and Buckel, Gene 51:31-41 (1987)). Clones providing for high levels of expression of t-PA are stored for subsequent use.

It is evident from the above results, that the subject method provides for a novel approach to expression of a wide variety of mammalian genes of interest. The method is simple, only requires the knowledge of a sequence of about 300 bp or more in the region of a target gene, and one may then use substantially conventional techniques for transferring the amplifiable region to an expression host, and production of the desired product in high yield.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made

thereto without departing from the spirit or scope of the appended claims.